Economics 102: Analysis of Economic Data
Cameron Spring 2015 May 14
Department of Economics, U.C.-Davis

Second Midterm Exam (Version A)

Compulsory. Closed book. Total of 30 points and worth 22.5% of course grade.
Read question carefully so you answer the question.
You are to use only **simple calculations (+, -, /, *, square root) and show all workings**.
For computations final answers should be to at least four significant digits.
You may remove the formula sheet and the Stata output sheet(s) at end of exam.

Question scores

Question	1*a*	1*b*	1*c*	1*d*	1*e*	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	3*a*	3*b*	3*c*	4*a*	4*b*	*Mult. choice*
Points	2	2	1	1	1	1	1	2	2	3	2	2	2	1	1	1	5

1.(a) Suppose a random sample of size 100 on variable x yields a sample mean of 50 and a sample standard deviation of 400. Give an approximate 95% confidence interval for the population mean.

(b) Provide the four population assumptions used for the linear regression model.
(0.5 points per correct assumption).

(c) Which of the assumptions given in part (b) are necessary for the OLS estimates to be unbiased?

(d) Suppose y has sample variance 1, x has sample variance 4, and the sample covariance between x and y is 1.
What is the sample correlation coefficient? Provide your calculations.

(e) Continuing with the previous question, what is the slope coefficient from regression of y on an intercept and x? Provide your calculations.

QUESTION 2 USES STATA OUTPUT GIVEN AT THE END OF THIS EXAM.
For some questions the answer is given directly in the output.
For other questions you will need to use the output plus additional computation.

The data are for state flagship universities in 2014

`Instate` = 2014-15 In-State Tuition and Fees in dollars

`OutofState` = 2014-15 Out-of-State Tuition and Fees in dollars

2.(a) How do out-of-state tuition and fees change when in-state tuition and fees increase by one thousand dollars?

(b) Give a **95 percent** confidence interval for the population slope coefficient.

(c) Give a **99 percent** confidence interval for the population slope coefficient.

(d) The claim is made that out-of-state tuition and fees are not associated with in-state tuition and fees. Test this claim at significance **level 0.05**. **State clearly the null and alternative hypotheses and your conclusion.**

(e) The claim is made that out-of-state tuition and fees increase by more than \$1 with each extra \$1 of in-state tuition and fees. Test this claim at significance **level 0.01**. **State clearly the null and alternative hypotheses and your conclusion.**

(f) UC Berkeley has in-state tuition and fees of \$12,972 and out-of-state tuition and fees of \$35,850. Are the out-of-state tuition and fees higher than expected given Berkeley's in-state tuition and fees? Explain your answer.

3. You are given the following for a data set on three observations

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	14	-2	-14	28	4	196
2	28	-1	0	0	1	0
6	42	3	14	42	9	196

(a) Calculate the OLS intercept and slope coefficients.

(b) Calculate the standard error of the regression.

(c) Calculate the R-squared of the regression.

4. You run the following code, where `runiform(-1,1)` draws random variables that have the uniform distribution on the interval -1 to 1 and have mean zero.

```
clear
set seed 10101
program myprogram, rclass
 drop _all
 quietly set obs 1000
 generate x = runiform(-1,1)
 generate u = runiform(-1,1)
 generate y = 1 + 2*x + u
 regress y x
 return scalar mystery = _b[x]
end
simulate mystery=r(mystery), seed(10101) reps(500): myprogram
summarize mystery
histogram mystery
```

(a) What do you expect the sample mean of variable `mystery` to be? **Explain.**

(b) What distribution do you expect variable `mystery` to have? **Explain.**

Multiple Choice Questions (1 point each)

1. For hypothesis testing

a. test size is the probability of a type I error

b. test power is one minus the probability of a type II error

c. both a. and b.

d. neither a. nor b.

2. For linear regression the conditional mean of y given $x = x^*$ equals

a. $b_1 + b_2 x^*$

b. $b_1 + b_2 x^* + e$

c. $\beta_1 + \beta_2 x^*$

d. $\beta_1 + \beta_2 x^* + u$

e. none of the above.

3. With a very large sample size and OLS estimation of a correctly specified regression model we can quite accurately predict

a. the conditional mean of y given $x = x^*$

b. the actual value of y given $x = x^*$

c. neither a. nor b.

d. both a. and b.

4. The OLS estimator

a. minimizes the sum of vertical deviations of actual data points from the regression line

b. minimizes the sum of horizontal deviations of actual data points from the regression line

c. both a. and b.

d. neither a. nor b.

5. The Stata command `regress y x, vce(robust)`

a. yields the same standard errors as command `regress y x`

b. yields the same t-statistics as command `regress y x`

c. both a. and b.

d. neither a. nor b.

SOME USEFUL FORMULAS

Univariate Data

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ *and* $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\bar{x} \pm t_{\alpha/2;n-1} \times (s_x/\sqrt{n})$ *and* $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

ttail$(df, t) = \Pr[T > t]$ where $T \sim t(df)$

$t_{\alpha/2}$ such that $\Pr[|T| > t_{\alpha/2}] = \alpha$ is calculated using invttail$(df, \alpha/2)$.

Bivariate Data

$r_{xy} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \dfrac{s_{xy}}{s_x \times s_y}$ [Here $s_{xx} = s_x^2$ and $s_{yy} = s_y^2$].

$\widehat{y} = b_1 + b_2 x_i$ $b_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $b_1 = \bar{y} - b_2\bar{x}$

TSS $= \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ ResidualSS $= \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ Explained SS = TSS - Residual SS

$R^2 = 1-$ ResidualSS/TSS

$b_2 \pm t_{\alpha/2;n-2} \times s_{b_2}$

$t = \dfrac{b_2 - \beta_{20}}{s_{b_2}}$ $s_{b_2}^2 = \dfrac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$

$y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$

$\mathrm{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

Multivariate Data

$\widehat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki}$

$R^2 = 1-$ ResidualSS/TSS $\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$

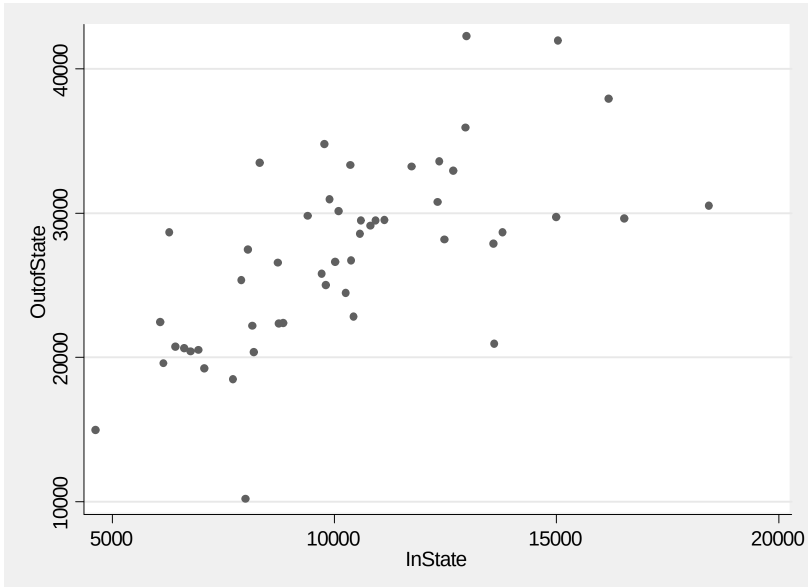
$b_j \pm t_{\alpha/2;n-k} \times s_{bj}$ *and* $t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$$

$$F = \frac{(\text{ResSS}_r - \text{ResSS}_u)/(k-g)}{\text{ResSS}_u/(n-k)} \sim F(k-g, n-k)$$

Ftail$(df1, df2, f) = \Pr[F > f]$ where F is $F(df1, df2)$ distributed.

F_α such that $\Pr[F > f_\alpha] = \alpha$ is calculated using invFtail$(df1, df2, \alpha)$.




```
. summarize InState OutofState

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     InState |         50     10293.3    3014.488       4646   18464.03
  OutofState |         50    27057.86    6398.589      10104      42184

. regress OutofState InState

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(1, 48)        =     34.58
       Model |   840017671         1   840017671   Prob > F        =    0.0000
    Residual |  1.1661e+09        48  24294527.3   R-squared       =    0.4187
-------------+----------------------------------   Adj R-squared   =    0.4066
       Total |  2.0062e+09        49  40941938.4   Root MSE        =    4928.9

------------------------------------------------------------------------------
  OutofState |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     InState |   1.373512   .2335837     5.88   0.000     .9038607    1.843164
       _cons |   12919.89   2503.352     5.16   0.000     7886.565    17953.22
------------------------------------------------------------------------------

. display "t_48,.005 = " invttail(48,.005) "  t_48,.01 = " invttail(48,.01)
t_48,.005 = 2.682204  t_48,.01 = 2.4065813

. display "t_48,.025 = " invttail(48,.025) "  t_48,.05 = " invttail(48,.05)
t_48,.025 = 2.0106348  t_48,.05 = 1.6772242
```